What will the scholarly profile page of the future look like? Provision of metadata is enabling experimentation.

blogs.lse.ac.uk/impactofsocialsciences/2015/07/16/scholarly-profile-of-the-future/

7/16/2015

From multi-stakeholder platforms like ORCID, to commercial services like Google Scholar, academic profiles exist in a complex landscape of information flows. **Lambert Heller** *provides an overview of the available scholarly profile pages and offers insight into their future development, which is set to be shaped by business models, technology, and available data streams.*

We're used to easily finding researchers' profile pages on the web, often finding more than one for the same researcher. Especially for the new generation of researchers, accustomed to having every piece of their research online, these individual profiles make all kinds of sense: They make it easy to show one's own efforts, maintain and present a network of related researchers, and reach out to collaborators and potential employers.

These pages are relatives of "professional networks" such as LinkedIn – but they offer many information elements and information flows that are specific to those working in the academy and scientific research. However, these profile pages go far beyond the digital version of a traditional CV ("curriculum vitae"). A workshop within the **Digital Humanities Experiments** event on 11/12 June 2015 at the German Historical Institute Paris (DHIP), led by mathematician David Chavalarias and me, explored tomorrow's networked researchers' profile pages. (Learn more about the workshop and its outcomes in my blog series around the event.)

We are not anywhere near a situation where one provider of scholarly profile pages makes all the others unnecessary. This is due to the fact that we have a complex landscape of information flows with a number of totally different information hubs.

It's about the data, stupid! – Why metadata availability largely defines the three major business models of scholarly profiles

On the one hand, we have contributor registries like ORCID, which is operated by a far-reaching multi-stakeholder coalition. They don`t even try to deliver one "full service package", but they want to provide a sustainable, agreed-upon source of author data available to all kinds of services. Then we have commercial services. Bundling all kinds of services on their own platform, these services are most often islands in regard to their preferred information flow. You fill out your profile, maybe even upload your papers, but in most cases you will have a hard time to let any other service reuse the data once you left it there. Third, we have institutional players, and some non-commercial players that are essentially financed by academic funding organizations. They often draw from publicly available information streams, like the before-mentioned ORCID – but if and how the resulting profile pages are fit for consumption on the public web depends heavily on the goals and means of the institution running the services.

Let's have a closer look at some of the major players in each of these categories. Or have a look at the table, if you're in hurry…

	Feature	ORCID	Research Gate (&	Mendeley	Pure (& similar CRIS)	VIVO	Google Scholar	Impact Story
	Business model	Multi-stakeholder	Commercial	Commercial	Academic institutions utilizing propreitary software& commercial services	Academic institutions utilizing Open Source software	Commercial	Academic 3rd party funding, user pays
Identity	Portrait picture	N	Y	Y	Y	Y	Y	Y
	Researchers alternative names	Y	N	N	N	N	N	N
	IDs / profiles in other systems	Y	N	N	N	N	N	Y
CV items	Papers and similar	Y	Y	Y	Y	Y	Y	Y
	Uncommon research products	Y	Y	Y	Y	Y	N	Y
	Grants, third party funding	Y	N	N	Y	Y	N	N
Institutional	Current institution	Y	Y	Y	Y	Y	Y	Y
	Former employers, education etc.	Y	Y	Y	Y	Y	N	N
Keywords	Self assigned keywords	Y	Y	Y	N	N	Y	N
	Concepts from controlled vocabulary	N	N	N	Y	Y	N	N
Social graphs	Social graph of followers/friends	N	Y	Y	N	N	N	N
	Social graph of coauthors	N	Y	Y	Y	Y	Y	N
Citation / attention data	Citation/attention metadata from platform itself	N	Y	Y	N	N	N	N
	Citation/attention metadata from other sources	N	N	N	Y	Y	Y	Y
Bundled functionality	Comprehensive search to match / include own papers	Y	Y	Y	N1	N2	Y	Y3
	Forums, Q&A etc.	N	Y	Y	N	N	N	N
	Deposit own papers	N	Y	Y	N	N	N	N
	Research administration tools	N	N	N	Y	Y	N	N
Data	Reuse of data from outside of the service	High	Low	High	Low	High	Low	High

[1] Often supplemented by institution
[2] Often supplemented by institution
[3] Via ORCID

What do networked researchers' profile pages include? Or, one ORCID iD to rule them all.

When we talk of modern approaches to the issue of networked profiles, we have to mention ORCID. ORCID is a relatively new initiative driven by some of the largest non-profit and commercial academic publishers, national libraries, professional societies and major Open Access repositories. Their goal is to build a centralised registry of all

“researchers and contributors” to academic products, allowing for unique identifiers that remove ambiguity regarding the identification of their contributions. As an example, take a look at the web representation of ORCID iD 0000-0001-5109-3700.

What does ORCID hope to achieve? First, all publishing and archiving outlets will sooner or later be able to identify all authors and contributors by their ID; second, institutions and individuals can populate their own profiles with the ORCID data collected about them, synchronising and updating between their ORCID profiles and any other profiles they may have elsewhere. But is there any need for other profiles if you can have everything in one place – ie your ORCID profile? Let’s explore this in greater detail …

Information elements:

- Scholarly products (articles for journal and other publications)
- Self-assigned keywords
- Researchers’ alternative names (to ensure disambiguation)
- Identities in other systems, profiles on other services
- Attribution of multiple institutions (education, former employers, etc.)
- Attribution of grants/third party funding

Reuse factor (structured availability and reuse rights):

- High

Kings of convenience: the rise of commercial siloed academic networks

While ORCID may be new to some readers, nearly everybody within or in the vicinity of the academic environment is now familiar with “Facebook for scientists” services such as ResearchGate. This type of service started gaining ground around 2008 – the leaders in the field being ResearchGate, academia.edu and Mendeley, with user counts allegedly in the millions. (For further analysis, cf.Nentwich and König 2014. Example of profile pages: ResearchGate, academia.edu, Mendeley.) One reason for their success must be the convenience they offer, enabling anyone to present their academic efforts in one place – a convenience that sometimes develops into rather aggressive urging of users to update their profile for better discoverability. A prime example of the strange outgrowth of this kind of service is the “ResearchGate score”, a self-acclaimed new measure for scholarly impact, an indicator based solely on activities occurring on this service’s website, possibly one of the purest offerings to scholarly vanity imaginable.

What all these Facebook-mimicking services have in common is that all of the information entered in the database of these services, from simple facts about a researcher’s work to whole papers that can be self-archived directly into these services, is owned solely by the commercial enterprises behind them. In this way, these services exemplify the “web 2.0” principle of being free (as in free beer), with the caveat that you cede control over your aggregated profile data. This is not only a matter of data-freedom principles. If you try to harvest large chunks of content from these databases for reuse elsewhere (as undertaken regularly by Google and other search engines), you soon learn that this is not permitted. Only Mendeley earns a special mention for being a kind of exception in this regard –it offers much of its data under reuse conditions.

Most common information elements:

- Scholarly products (articles for journal and other publications)
- Self-assigned keywords
- Attribution of multiple institutions (education, former employers, etc.)

- Personal profile photo
- Social graph (type of follower relation, in some services co-authorship)
- Attention metadata from the platform itself (views, downloads, bookmarks, etc.)

Reuse factor (structured availability and reuse rights):

- Low to non-existent (most academic networks)
- High (Mendeley)

Authentic researcher profiles that are (almost) never meant for the public web: siloed institutional "current research information systems" (CRIS)

Although information systems such as ResearchGate tend to be very popular at present, and can by all means shed light on what scholars truly want ™, they have at least one enduring problem: they are never complete. However, if you define scholarship as being attached to a certain university or other research institution, you may find "current research information systems" (CRIS) to be a possible new contender for acting as a valuable source of information about researchers and their activities. And a complete one at that, at least with regard to the institution running the respective CRIS.

What are CRIS all about? Mainly acquired by large academic publishers in recent years, contenders such as Thomson Reuters Converis, Elsevier Pure and Symplectic Elements offer CRIS database products. Research institutions run CRIS to pool data about their staff and research facilities. From a research controlling perspective, this is useful for understanding and reinforcing an institution's assets. Although most of these systems are, technically, online databases, only a few institutions view this as an opportunity to raise public awareness of their research activities. In many cases, databases are completely hidden from public view. In contrast to "Facebook for scientists" services of the ResearchGate kind, with CRIS we have no problems with completeness and re-usage rights, but with the public availability of the data in the first place. That said, there are a number of positive exceptions: as mentioned in an earlier blog post, VIVO aims to be a research information system based on the original means of the web (like semantic ontologies), while delivering information from some universities to the whole open web, usually including comprehensive re-usage rights.

Most common information elements:

- Scholarly products (articles for journal and other publications)
- Detailed attribution of institutional roles and positions
- Self-assigned keywords
- Concepts from controlled vocabularies and/or automatically generated profiles
- Personal profile photo
- Social graph (co-authorship)
- Attribution of grants/third party funding

Reuse factor (structured availability and reuse rights):

- Low to non-existent (most CRIS implementations)
- High (VIVO, a number of other CRIS implementations)

Impact and other ways to tell a scholar's story: other approaches to researcher profile pages

Another very well-known type of researcher profile pages is delivered by Google's academic search engine

“Scholar”. (According to preliminary results of a very interesting survey from Utrecht University library, GS profiles are even more popular than those on ResearchGate, let alone ORCID, academia.edu or institutional CRIS.)
Google Scholar is more or less comparable with huge traditional science citation indexes such as Web of Science (or WoS for short, now owned by Thomson Reuters) and its rival, Elsevier Scopus. Google radicalised competition between these huge cross-disciplinary corpora of scientific article and citation metadata: while WoS and Scopus covered a limited set of peer-reviewed academic journals, placing them all in an online database licensed by university libraries, Google Scholar takes a full-text search engine approach, undeniably covering more documents and delivering search results, including citation counts, to end users for free.

In 2011, Google Scholar launched profiles, something that cannot be found in WoS or Scopus. The idea is not only to give searchers a comprehensive view of individual researchers, their articles and citation counts, but also to enable them to add to their profiles themselves. Unlike ResearchGate, the service does not aim to be a “full service package”. Instead of inviting researchers to self-archive their papers on the actual website, it covers self-archived versions from services such as ResearchGate as well as from traditional institutional Open Access repositories. The only data that Google Scholars may automatically give to third party services is the citation count of each document.

ImpactStory offers a service that is comparable in many ways to that of Google Scholar profiles. However, it follows a very different business model. (Example of an ImpactStory profile page.) While Google Scholar is a commercial service that searchers and profile owners can use free of charge, ImpactStory is a largely third party-funded non-profit organisation seeking to become sustainable through services paid for by profile owners. While Google Scholar draws its data from its own article and citation index, ImpactStory remains sleek by drawing from many different sources of citation data and impact metadata – from Facebook ‘likes’ to the number of forks on Github – or so-called “altmetrics”. The idea is to operate as a service for collecting and consolidating this data, and to present it on behalf of profile owners.

ImpactStory is by no means the only service that aspires to be the clearing point for this kind of data – compare, for example, Plum or Altmetric.com. In the growing landscape of citation and attention metadata, many publishers, repositories and institutional research information services have already decided not to collect impact metadata themselves, but to draw from one of these services. It is interesting to note that ImpactStory was one of the first services of its kind to offer the automated import of ORCID data. To conclude: although they appear to be similar at first glance, Google Scholar profiles are a strongly shielded island, whereas ImpactStory strives to be a useful intersection for different services and data streams.

Common information elements:

- Scholarly products (articles for journal and other publications)
- Self-assigned keywords
- Personal profile photo
- Social graph of co-authorship (Google Scholar)
- Social graph (type of follower relation, in some services co-authorship)
- Citation generated on the platform itself (Google Scholar)
- Citation and other impact data from different platforms (ImpactStory)

Reuse factor (structured availability and reuse rights):

- Low to non-existent (Google Scholar)
- High (ImpactStory)

Some conclusions

With the growing expectations of cultivating one's own scholarship profile online completely **and** conveniently, things have become more interesting, and sometimes confusing. The whole area still seems to be in its infancy. A strong indicator of the ongoing development of this ecosystem is the consolidation of freely available metadata streams – besides ORCID, we now have CrossRef's DOI event tracker pilot as a free source of impact metadata across many scholarly articles. In the area of institutional research information systems, open approaches such as VIVO ontologies and software are constantly gaining greater traction, enabling custom developments and experimentation. So, interesting times ahead!

Disclaimer: On behalf of my employer, TIB Hannover, I work with the DOI event tracker working group and the TIB Open Science Lab runs experiments and development with VIVO ontologies and software.

*This topic was covered at a workshop within the **Digital Humanities Experiments** event on 11/12 June 2015 at the German Historical Institute Paris (DHIP), led by mathematician David Chavalarias and me. This is an edited extract of a piece which originally appeared here and is the second part (read part 1) of a contribution to DHIP's blog carnival accompanying the whole event.*

Note: This article gives the views of the authors, and not the position of the Impact of Social Science blog, nor of the London School of Economics. Please review our Comments Policy if you have any concerns on posting a comment below.

About the Author

Lambert Heller *serves currently as the head of Open Science Lab at TIB Hannover (German national library of science and technology). As an academic librarian with a background in humanities and social sciences, he tries to find useful new things in the area of scholarly communication, and he writes and teaches sometimes about that. He tweets as @Lambo.*